Enhancing Marine Habitats Detection: A Comparative Study of Semi-Supervised Learning Methods

Rim Rahali[1], Thanh Phuong Nguyen[2] and Vincent Nguyen[1]
[1]*University of Orleans, INSA-CVL, LIFO UR 4022, Orleans, France*
[2]*I3S, CNRS, UMR 7271, University of Cote d'Azur, France*

Keywords: Semi-Supervised Learning, UIE Methods, Underwater Images, Object Detection, Contrastive Learning.

Abstract: Most of the recent success in applying deep learning techniques to object detection relies on large amounts of carefully annotated and large training data, whereas annotating underwater images is a costly process and providing a large dataset is not always affordable. In this paper, we conduct a comprehensive analysis of multiple semi-supervised learning models, used for marine habitats detection, aiming to reduce the reliance on extensive labeled data while maintaining high accuracy in challenging underwater environments. Results, performed on Deepfish and UTDAC2020 datasets attest a significant performance conducted by semi-supervised learning, in terms of quantitative and qualitative evaluation. An other study related to Underwater Image Enhancement (UIE) methods and contrastive learning is presented in this work to deal with underwater images specificity and provide more comprehensive analysis of their impact on marine habitats detection.

1 INTRODUCTION

Detecting marine habitats, or more broadly, Underwater Object Detection (UOD) represents a challenging research topic, where difficult underwater environments make underwater images suffer from noise, blur, low contrast, diffusion effect and color distortion (Sarkar et al., 2022). Various UOD techniques based on deep learning were developed in this context, helping researchers to reach to new levels in exploring the underwater world (Han et al., 2020; Pan et al., 2021). Although important results that have been achieved over the years, UOD techniques are still limited in front of: 1) The insufficiency in underwater image dataset, 2) The low quality of images due to complex underwater environment, and 3) The large number of required labeled images while it is expensive to annotate and acquire them. Most of deep learning algorithms rely on the availability of large, well-balanced and labelled datasets. This type of supervised pipeline can not handle the specificities of underwater imaging.

Semi-Supervised Object Detection (SSOD) (Wang et al., 2023) has become an active task in recent years to deal with label expenditure. It uses both labeled data and unlabeled data for training where unlabeled data are more explored for boosting object detectors and they are relatively easy to collect. The challenge remains in how to use effectively these unlabeled data. Teacher-student learning models were widely used for SSOD (Mi et al., 2022; Li et al., 2023) and achieved notable success. They consist of two networks: 1) The teacher network to generate pseudo-labels for unlabeled data, and 2) The student network to be trained using both the generated pseudo-labels and ground truth. The student model updates its weights by training, and the teacher updates its weights from the student model by Exponential Moving Average (EMA) (Tarvainen and Valpola, 2017). Besides, strong and weak data augmentations are separately applied to enforce the consistency between the two networks (Cubuk et al., 2019; Xie et al., 2020). While SSOD methods can exploit large amounts of unlabeled data to address the issue of insufficient labeled data in UOD, they has not gained enough attention in the field of underwater applications and existing works are still limited (Zhou et al., 2023). The complexity and diversity of underwater environments, characterized by low contrast, blur, color distortion, hazing, and more; introduce additional difficulties that make UOD more challenging than general object detection.

To the end, we propose a comprehensive analysis of performance using semi-supervised models, applied to different marine habitats datasets. On the other hand, we analyse the impact of Underwater Im-

Rahali, R., Nguyen, T. P. and Nguyen, V.
Enhancing Marine Habitats Detection: A Comparative Study of Semi-Supervised Learning Methods.
DOI: 10.5220/0013325500003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 233-244
ISBN: 978-989-758-728-3; ISSN: 2184-4321

age Enhancement (UIE) methods on the performance of these models. UIE methods are widely applied to remove blurring, color distortion in images, improving the features of interesting targets while reducing those of irrelevant background (Xu et al., 2023). In addition, we integrate contrastive learning (Zhang et al., 2022b) into existing SSOD methods. It is an approach that aims to minimize the distance between similar data points while maximizing the distance between dissimilar ones in the embedding space. Implementing contrastive learning can lead to improved feature learning and better overall detection capabilities of underwater object detectors. The main contributions of this work can be listed as follows.

1. We conduct a comprehension analysis of performance of three popular SSOD methods, Active Teacher (Mi et al., 2022), Unbiased Teacher (Liu et al., 2021), Robust Teacher (Li et al., 2023) on two marine habitats datasets: Deepfish (Saleh et al., 2020) and UTDAC2020 (Song et al., 2023).
2. We evaluate different UIE methods applied to Deepfish and UTDAC2020 datasets and analyse their impact on detecting marine habitats using semi-supervised methods.
3. We incorporate contrastive learning into semi-supervised models and evaluate its impact on detection for Deepfish and UTDAC2020 datasets.

This paper is organized as follows. Section 2 illustrates related works; Section 3 presents preliminaries; we exhibit details of our methodology in Section 4; we present the experimental results and analyses in Section 5; and, finally, Section 6 concludes the paper.

2 RELATED WORK

2.1 Underwater Object Detection

In recent years, research on underwater object detection has undergone a notable transformation, moving from the use of traditional manual features to embracing deep learning techniques. Initially, traditional manual features were used in early stages of research (Yu, 2020). However, these approaches face significant limitations when applied to practical underwater environments. Furthermore, most of underwater object detection algorithms that rely on manual feature extraction process, require professional expertise and complex algorithm debugging. Recently, the development of machine learning has contributed to undergoing research dedicated to underwater object detection. Methods developed in this field involve extracting and combining traditional artificial features, such as texture, shape, color and target movement, and then using them in conjunction with machine learning algorithms to perform underwater object detection. For example, in (Srividhya and Ramya, 2017), the authors proposed a strategy that combines learning algorithms with texture features for accurate detection and recognition of underwater objects. Here, the texture features are valuable indicators of the surface properties of an image and they play a significant role in different underwater detection scenarios. In addition to texture, color and motion features play a major role in the analysis of underwater images. These have been studied in different works. For example, the authors in (Chen and Chen, 2010) proposed a new color edge detection algorithm that uses the Kuwahara filter (Bartyzel, 2016) to smooth the original image. They have integrated adaptive thresholding and contour spacing algorithms to improve detection efficiency and performance.

Recently, new methods based on deep learning have become increasingly important for their ability to automatically learn and extract features from underwater images. This can replace underwater object detection methods that rely on manual feature extraction. In (Han et al., 2020) researchers combined max-RGB and grayscale methods to boost underwater vision. Then, by obtaining illumination maps, they introduced a CNN method to solve the problem of low illumination in underwater images. Similar, in (Chen et al., 2020), the authors developed an architecture called Sample Weighted hypernetwork (SWIPENet) for detecting small underwater objects. The architecture improve the accuracy of object detection, dealing with the image blur. Numerous object detection algorithms marked a pivotal moment in the rapid progress of deep learning in underwater object detection. For example, an enhanced YOLOv5 algorithm was proposed in (Ren et al., 2022) specifically for underwater object detection. The authors incorporated the twin transformer as the backbone network and improved the multiscale feature fusion method and confidence loss function. In (Lau and Lai, 2021), the authors focused on the selection and enhancement of the basic network architecture in Faster R-CNN. They performed pre-processing on the obtained images and tested the performance of different network architectures to identify the most suitable one for training object detection in turbid media. Furthermore, to deal with the limited underwater image data that impact the prediction results, an unsupervised knowledge transfer (UnKnoT) was introduced, in (Zurowietz and Nattkemper, 2020). The method uses a data augmentation technique, called scale transfer to reuse existing training data and detect the same object classes in a

new image dataset.

2.2 Semi-Supervised Underwater Object Detection

In underwater object detection tasks, the limited amount of underwater image data poses a significant challenge. In response, researchers have adopted semi-supervised approaches to address this problem and improve the detection capability of underwater object detection algorithms. In (Jahanbakht et al., 2023), a two phase semi-supervised contrastive learning approach was developed to reduce the impact of reliance on a high volume of accurately labeled data. The proposed model consists of a self-supervised contrastive learning phase, followed by fully-supervised incremental fine tuning learning to detect various fishes in turbid underwater video frames. A teacher-student model was proposed in (Alaba et al., 2023) to recognize fish species. The teacher network generates pseudo-labels, and the student network is trained with the generated pseudo-labels and ground truth simultaneously. The model consists of a Faster R-CNN with Feature Pyramid Network (FPN) detector. In (Zhou et al., 2023) an novel underwater object detection framework, named UWYOLOX, was presented as joint learning-based underwater image enhancement module (JLUIE) and an improved semi-supervised learning method USTAC. JLUIE and YOLOX-Nano (Ge et al., 2021) share the detection loss for training, where JLUIE can adaptively enhance each image for better detection performance. Then, USTAC is introduced to further improve the mean Average Precision of object detection.

Although semi-supervised learning has a relatively long history, it has only recently gained widespread attention in underwater domain applications. Ongoing research is focused on better understanding the underwater environment and incorporating its specific features into semi-supervised models, with the aim of improving the effectiveness of these approaches in such challenging conditions. The focus of this work is to adapt general semi-supervised learning methods, particularly teacher-student models, to the domain of underwater imaging. To achieve this, we conduct a comprehensive analysis of marine habitats detection, performed using popular SSOD methods: Active Teacher, Unbiased Teacher, and Robust Teacher. These methods, applied for the first time to the Deepfish (Saleh et al., 2020) and UTDAC2020 (Song et al., 2023) datasets, were chosen for their popularity and their ability to represent diverse strategies within teacher-student architectures. While they are not the current SOTA in SSOD, they remain highly influential in the field, making them ideal candidates for a comparative study that aims to highlight the strengths and weaknesses of different SSOD methods.

3 PRELIMINARIES

In this section, we present three popular semi-supervised methods, used in literature for object detection tasks. They share the principle of based teacher-student mutual learning, which is a common approach used to train models with limited labeled data and a larger amount of unlabeled data. While Teacher and Student are given weakly and strongly augmented data as inputs, respectively, the Teacher network is responsible for generating pseudo-labels for unlabeled data, and the student will be trained using both pseudo-labels and ground truth (of labeled data). At this stage, the student incorporate consistency regularization techniques (Jeong et al., 2021) to ensure its robustness at producing the outputs although the presence of small perturbations. Besides, the teacher's weights θ_t are updated during the semi-supervised training by EMA (Tarvainen and Valpola, 2017) of the student's weights θ_s:

$$\theta_t^i \leftarrow \alpha\theta_t^{i-1} + (1-\alpha)\,\theta_s^i \tag{1}$$

, where i denotes the i^{th} training step and α determines the speed of the transmission. The weights of student network θ_s are updated using back propagation. The model's optimization process is formulated as minimizing the loss L:

$$L = \lambda_s L_{sup} + \lambda_u L_{unsup} \tag{2}$$

, where L_{sup} and L_{unsup} represent the supervised and the unsupervised losses respectively. λ_s and λ_u are pondering coefficients for L_{sup} and L_{unsup}, respectively.

3.1 Unbiased Teacher

The main idea of Unbiased Teacher (Liu et al., 2021) is to introduce a class-balance Focal Loss (Zhang et al., 2022a) to address the pseudo-labeling bias issues caused by class-imbalance existing in ground truth labels. Besides, to minimize the bias, the Unbiased Teacher uses a novel data augmentation technique called BoxJitter which is applied to make the student more robust toward object localization and helps reduce localization bias in pseudo-labels. In the other hand, a high filtering threshold is used for pseudo-labels to ensure that only high-quality pseudo-labels are used for training, and the teacher

do not misguide the student. The presence of noisy pseudo-labels can affect the pseudo-label generation model. As result, the Teacher and the student are detached, only the learnable weights of the Student model is updated via back-propagation by using a supervised loss L_{sup} and a unsupervised loss L_{unsup}. Given a set of labeled data $D_L = \{X_L, Y_L\}$ and a set of unlabeled data $D_U = \{X_U\}$, where X denotes the data and Y is the label set. X_L, Y_L, and X_U are defined as $X_L = \{x_i^l, i \in N_l\}$, $Y_L = \{y_i^l, i \in N_l\}$, and $X_U = \{x_i^u, i \in N_u\}$, respectively where N_l represents the number of labeled examples and N_u the unlabeled ones. For the Unbiased Teacher, the loss is composed of the supervised loss L_{sup} and the unsupervised loss L_{unsup}, defined as:

$$L_{sup} = \frac{1}{N_l}\sum_{i=1}^{N_l}\Big(L_{cls}^{rpn}(x_i^l, y_i^l) + L_{reg}^{rpn}(x_i^l, y_i^l) + L_{cls}^{roi}(x_i^l, y_i^l) + L_{reg}^{roi}(x_i^l, y_i^l)\Big) \tag{3}$$

$$L_{unsup} = \frac{1}{N_u}\sum_{i=1}^{N_u}\big(L_{cls}^{rpn}(x_i^u, \hat{y}_i^u) + L_{cls}^{roi}(x_i^u, \hat{y}_i^u)\big) \tag{4}$$

, where, L_{cls}^{rpn}, L_{reg}^{rpn}, L_{cls}^{roi}, L_{reg}^{roi} represent the Region Proposal Network (RPN) classification loss, the RPN regression loss, the Region of Interest (ROI) classification loss, and the ROI regression loss respectively. Here, $\hat{y}_i$ represent the generated pseudo-label.

3.2 Robust Teacher

The main focus of the Robust Teacher (Li et al., 2023) is to address the noisy labels. The Robust Teacher dealt with this challenge from two perspectives: 1) Developing a vise Self-Correcting Pseudo-labels Module (SPM) to addresses noise in pseudo-labels by refining object localization first and then improving class predictions, reducing errors in both, and 2) Mitigating the inherent class bias in pseudo-labels by introducing the Re-balanced Focal Loss (FL) which adjusts the loss function to focus more on under-represented classes, preventing the model from being biased toward dominant classes. Together, the Robust Teacher ensures that the pseudo-labels used for training are both more accurate and better balanced across different object classes. The loss function is summarized as the sum of the supervised loss L_{sup} and the unsupervised loss L_{unsup}, described as:

$$L_{sup} = \frac{1}{N_l}\sum_{i=1}^{N_l}\Big(L_{cls}^{rpn}(x_i^l, y_i^l) + L_{reg}^{rpn}(x_i^l, y_i^l) + L_{cls}^{roi}(x_i^l, y_i^l) + L_{reg}^{roi}(x_i^l, y_i^l) + L_{cls}^{ml}(x_i^l, \nu_i^l)\Big) \tag{5}$$

$$L_{unsup} = \frac{1}{N_u}\sum_{i=1}^{N_u}\big(L_{cls}^{rpn}(x_i^u, \hat{y}_i^u) + L_{cls}^{roi}(x_i^u, \hat{y}_i^u) + L_{cls}^{ml}(x_i^u, \nu_i^u)\big) \tag{6}$$

L_{cls}^{ml} is the Multi-Label (ML) head classification loss (Zhang et al., 2022a). In fact, a ML head was introduced into the Faster-RCNN detector to predict image-level pseudo-labels ν_i for class distribution re-balancing to alleviate the inherent class imbalance issues. The ML head takes the top-level feature of Feature Pyramid Network (FPN) as inputs and uses the sigmoid function to convert the output into a multi-label probability distribution which used to calculate a re-balanced weight w for the re-balanced focal loss L_{cls}^{RFL} given as:

$$L_{cls}^{RFL} = wy^T L_{cls}^{FL} \tag{7}$$

, with y_i and L_{cls}^{FL} represent the category label and the focal loss, respectively. Here, L_{cls}^{ml} integrates the contribution of L_{cls}^{RFL} in the handling of rare classes and the refinement of classification.

3.3 Active Teacher

The Active Teacher (Mi et al., 2022) is characterized by its active learning, where the label set is partially initialized and gradually augmented by evaluating three key metrics of unlabeled examples: Difficulty, Information, and Diversity, used in combined manner (Cho et al., 2022). The method aims to improve the learning by selecting the most informative unlabeled data to label. Therefore, the Active Teacher can achieve high accuracy detection with fewer label set. Here, the supervised loss L_{sup} is defined as:

$$L_{sup} = \frac{1}{N_l}\sum_{i=1}^{N_l}\Big(L_{cls}^{rpn}(x_i^l, y_i^l) + L_{cls}^{roi}(x_i^l, y_i^l) + L_{loc}(x_i^l, y_i^l)\Big) \tag{8}$$

, with

$$L_{loc}(x_i^l, y_i^l) = \sum_{c\in\{x,y,h,w\}} \text{Smooth}_{L_1}(t_i^c - y_i^c) \tag{9}$$

and the unsupervised one L_{unsup} is defined as Eq.(4). L_{sup} consists of the classification loss L_{cls} of RPN and ROI head, and the one for bounding box regression L_{loc}. It is defined as the summation of the classification loss which presents the log loss over two classes (object vs. not object) and the bounding box regression loss. Here, t^c is the c^{th} coordinate of the output image x_i. L_{unsup} uses only the pseudo-labels of RPN and ROI head predictions. This loss is not applied for the bounding box regression since the confidence thresholding is not able to filter the pseudo-labels that are potentially incorrect for bounding box regression. The confidence of predicted bounding

boxes only indicate the confidence of predicted object categories instead of the quality of bounding box locations (Jiang et al., 2018).

4 PROPOSED METHODOLOGY

In this work, we propose a comprehensive analysis of the performance of different SSOD methods applied to marine habitats detection. To effectively apply SSOD methods, we propose the integration of two key modules for improving performance: the Underwater Image Enhancement (UIE) and the contrastive learning. The UIE is designed to address the challenges posed by underwater environments, such as color distortion, low contrast, and hazing, by enhancing the quality of the input images before they are processed by the model. We explore various UIE methods to improve image clarity, color balance, and detail sharpness. In addition to image enhancement, we introduce a contrastive learning strategy, which is integrated into the SSOD framework to help the model better differentiate between objects. In the following, first, we detail the different UIE methods, and second the contrastive learning strategy for marine habitats detection.

4.1 Underwater Image Enhancement

Underwater image enhancement methods are proposed to improve the visual quality of images captured underwater, which may suffer from hazing, low contrast, and color distortion/dominance. These methods were investigated with the aim of integrating UOD methods to achieve enhanced results. For the same reason, we investigate UIE for the SSOD methods. In the following sections, we present three distinct UIE techniques among the techniques analysed in (Ancuti et al., 2017; Islam et al., 2020; Song et al., 2020; Peng et al., 2023; Zhou et al., 2023), that achieve the highest UIQM and UCIQE scores (Xu et al., 2023) on Deepfish and UTDAC2020 datasets. UIQM and UCIQE are widely used metrics to assess the quality of enhanced images and evaluate UIE methods.

4.1.1 UIE-1: Color Balance and Fusion

The method[1] is based on color balance and fusion to enhance the image clarity and corrects the color distortion (Ancuti et al., 2017). The color balance composed helps to correct the color cast by adjusting the color channels so that their averages are equal. Then, a multi-scale fusion technique (Ancuti et al., 2012) is applied to combine several enhanced versions of the image that is directly derived from the color balanced version of the original degraded image; Each image is optimized for specific characteristics such as contrast and detail. This fusion uses weight maps to select the sharpest and most contrasted parts of each version, resulting in a final image that is more balanced, with natural colors, improved contrast, and sharper details. The first input of the the fusion process is a gamma corrected image of the white balanced image version, that aims to correct the global contrast. This correction increases the difference between darker/lighter regions at the cost of a loss of details in the underexposed regions. To compensate for this loss, a second input is generated, corresponds to a sharpened version of the white balanced image. A normalized unsharp masking process is applied with:

$$S = (I + N\{I - G * I\})/2 \tag{10}$$

, where I is the white balanced image, $G * I$ denotes the Gaussian filtered version of I. N. represents the linear normalization operator, also named histogram stretching in the literature. This operator shifts and scales all the color pixel intensities of an image with a unique shifting and scaling factor defined so that the set of transformed pixel values cover the entire available dynamic range.

4.1.2 UIE-2: U-Shape Transformer

The U-Shape Transformer[2] is a deep learning network (Peng et al., 2023), which combines the strengths of the U-Net and Transformer models, to ensure color correction, visibility improvement, and artifact reduction. Inspired by U-Net, the U-shape structure is designed to capture multi-scale information through an encoder-decoder architecture. The encoder downsamples images to extract high-level features, while the decoder up-samples to restore image resolution. Transformer blocks are integrated into both the encoder and decoder to capture long-range dependencies and global context, helping the model manage spatial complexity and variations, especially in underwater images. Skip connections between the encoder and decoder merge local and global features, leading to more accurate image enhancement.

The U-shape Transformer includes two specialized modules, based generator and discriminator: A Channel-wise Multi Scale Feature Fusion Transformer (CMSFFT), and a Spatial-wise Global Feature Modeling Transformer (SGFMT) (Peng et al., 2023). The SGFMT was designed, based on the spatial self-attention mechanism to replace the original

[1] https://github.com/Sai-paleti25

[2] https://github.com/LintaoPeng

bottleneck layer of the generator. It can accurately model the global characteristics of underwater images and reinforce the network's focus on the space areas with more serious attenuation, thus achieving uniform UIE. The CMSFFT module is responsible for processing features across different channels and scales. It replaces the skip connection of the generator and employs a channel-wise self-attention mechanism. This mechanism performs channel-wise multi-scale feature fusion on the features output by the generator's encoder. The fusion results are then transmitted to the decoder, reinforcing the network's attention to the color channels that experience more serious attenuation.

4.1.3 UIE-3: JLUIE Module

A joint learning-based underwater image enhancement module (JLUIE) was proposed in (Zhou et al., 2023), where four enhancement filters are applied in sequence. The White balance, Gamma correction, Contrast adjustment, and Sharpen contribute differently to image enhancement as follows: First, the White balance adjusts the colors of an image by calibrating the intensities of the red, green and blue channels to neutralize any color cast and make white objects appear white in the image. With $P_i = (r_i, g_i, b_i)$ the value of input pixel, the mapping function is :

$$P_o = (W_r r_i, W_g g_i, W_b b_i) \tag{11}$$

, where $P_o = (r_o, g_o, b_o)$ is the value of output pixel, (r,g,b) represent the red, green, and blue color channels respectively. W_r, W_g , W_b are the coefficients of the three color channels of red, green and blue respectively. Next, the mapping function of the Gamma correction filter is applied as $P_o = P_i^G$ with G is the Gamma value. The latter affects the overall brightness and contrast of the image. Then, a contrast adjustment is applied to modify the distribution of brightness levels in the image. This process enhances light areas, making them brighter, while dark areas become darker, using this mapping function:

$$P_o = \alpha En(P_i) + (1-\alpha)P_i \tag{12}$$

, where $En(P_i)$ represents the enhanced pixel value and α is a linear interpolation between the original image and the enhanced image. The last filter to apply is the Sharpen. It is used to remove image blur and sharpen contours and objects, using the following mapping filter :

$$F = I + \lambda(I - Gau(I)) \tag{13}$$

, where I and F are the input and output images respectively, Gau(I) denotes the result of applying a Gaussian filter to the input image, and λ is a positive scale factor. For this work, we use our proper implementation of JLUIE module.

4.2 Contrastive Learning

The main idea is to introduce a contrastive learning branch to the semi-supervised model to optimize pseudo-labels prediction based on the principle of pulling similar images together and pushing away the dissimilar ones. We couple the contrastive learning with the teacher-student architecture used in SSOD via the loss optimization. A new loss is added to the supervised loss L_{sup} and the unsupervised loss L_{unsup}, called the contrastive loss L_{ctr}.

$$L = \lambda_s L_{sup} + \lambda_u L_{unsup} + \beta L_{ctr} \tag{14}$$

, where β present the pondering coefficient for L_{ctr}. Similar to that in (Zhang et al., 2022b), it is formulated as:

$$\begin{aligned} L_{ctr} &= -\log\left(\frac{\sum_{k^+} exp(\gamma(\alpha_p s_p - m))}{\sum_{k^+} exp(\gamma(\alpha_p s_p - m)) + \sum_{k^-} exp(\gamma s_n)}\right) \\ &= \log\left(1 + \sum_{k^-} exp(\gamma(s_n + m)) \sum_{k^+} exp(-\gamma\alpha_p s_p)\right) \end{aligned} \tag{15}$$

Here, s_p represents the similarity of positive samples while s_n represents the similarity of negative samples. α_p, γ, and m are the soften parameter, the scale and the margin value (Zhang et al., 2022b), respectively. The similarity of positive and negative samples are averaged using the cosine distance (Popat et al., 2017), defined as:

$$s_{i,j} = \frac{x_i \cdot x_j}{\|x_i\| \left\|x_j\right\|} \tag{16}$$

, where, $x_i \cdot x_j$ represents the dot product between two sample vectors x_i and x_j, $\|x_i\|$ and $\left\|x_j\right\|$ represent their norms, respectively. An effective sampling strategy for positive and negative examples is crucial in non-supservised contrastive learning. In our method, we leverage the abundance of unlabeled data and the pseudo-labels generated by SSOD frameworks to select positive and negative samples. We expect that the co-optimization of the pseudo-labels generation alongside the contrastive loss helps improve the quality of pseudo-labels and the diversity of samples, which in turn enhances the learnt representations, leading to better overall detection performance. The unlabeled example x_i^u with pseudo label $\hat{y}_i^u$ is assigned to the most corresponding class c. Then, all the samples that have the same class c are pulled together, sharing the same specific instances corresponding to that class. In this way, the positives samples are created while the negatives are the samples that are pushed away with different class. With the contrastive branch, more meaningful representations are extracted which are involved in generating more reliable pseudo-labels.

5 EXPERIMENT

5.1 Datasets and Metrics

5.1.1 Datasets

We perform rigorous experiments on the challenging marine habitats datasets UTDAC2020 and Deepfish to evaluate the generalization performance of our approach. These datasets are specifically selected for their complexity and variability, providing a robust framework for testing the efficacy of the SSOD approach in diverse underwater scenarios.

DeepFish Dataset: DeepFish (Saleh et al., 2020) is a large-scale marine habitats dataset consisting of around 40 thousand images obtained from 20 different marine habitats in tropical Australia. Each habitat is divided into images with no fish (background) and images with at least one fish (foreground). The dataset is split into 50% training, 20% validation, and 30% testing, ensuring equal numbers of background and foreground images across all splits. All annotations are provided.

UTDAC2020 Dataset: UTDAC2020 (Song et al., 2023) is an underwater dataset derived from the underwater target detection algorithm competition 2020. There are 5168 training images and 1293 testing images. It contains four classes: echinus, holothurian, starfish, and scallop.

5.1.2 Metrics

We evaluate the semi-supervised models against the Average Precision (AP) (Sohn et al., 2020). It is a standard metric for object detection that measures the overlap between the prediction and the ground truth with Intersection Over Union (IOU) threshold set from 0.5 to 0.95, with 0.05 as the interval. The AP is calculated as:

$$AP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{17}$$

In marine habitat detection, a key challenge lies in accurately identifying and classifying habitats that often appear as small or medium-sized objects within images. Given the limited spatial area of the paper, in our experiments, we focus on three AP metrics described in Table.1 to analyse SSOD methods.

Table 1: The AP metrics used in our experiments.

Metrics	Description
AP	The mAP (mean average precision)
AP_S	The AP of small targets
AP_M	The AP of medium targets

5.2 Settings and Implementation Details

5.2.1 Experimental Settings

We propose to evaluate the performance of Active Teacher, Unbiased Teacher, and Robust Teacher on two different underwater datasets: Deepfish and UTDAC2020 datasets. Additional results are presented, investigating the performance of these methods from two aspects: 1) Applying various UIE methods to enhance input data, 2) Incorporating a contrastive loss into semi-supervised models to improve representation learning. Faster-RCNN (Ren et al., 2015) is defined as our supervised baseline for comparison with the semi-supervised methods analysed in our work.

Specifically, we use UTDAC2020 and Deepfish datasets to examine the SSOD methods on different experimental scenarios. In our setup, we randomly sample 40% labeled training data as our labeled set, with the remaining data serving as the unlabeled set. Unless stated otherwise, all tables present the results of models trained using the same 40% labeled data.

5.2.2 Implementation Details

Our implementation follows existing state of the art works (Mi et al., 2022; Li et al., 2023) and thus, Faster R-CNN is used with FPN and ResNet-50 backbone (He et al., 2016) as the default detector in the semi-supervised frameworks. Besides, ImageNet pretrained weights are used to initialize the feature extraction networks. We used SGD optimizer with the learning rate equals to 0.02 and momentum rate equals to 0.9. The supervised, unsupervised, and contrastive loss weights are equals to $\lambda_s = 0.5$ and $\lambda_u = 4.0$, and $\beta = 5.0$ respectively. We set $\alpha = 0 : 9996$ for EMA. We use confidence threshold $\tau = 0.7$ to filter the pseudo-labels of low quality. For the contrastive branch, we set $\alpha_p = 4$, m = 1, and $\gamma = 2$. The total training steps for each semi-supervised learning are 18000. In training, the unlabeled and labeled data are combined in the same proportion via random sampling, to create a mini-batch of size 20 which includes 10 labeled images and 10 unlabeled images.

For the data augmentation, we apply random horizontal flip for weak augmentation and randomly add color jittering, grayscale, Gaussian blur, and cutout patches for strong augmentations. This configuration is common on all three SSOD methods (Mi et al., 2022; Li et al., 2023; Liu et al., 2021).

(a) Robust Teacher (b) Unbiased Teacher (c) Active Teacher

Figure 1: Rows 1 and 2 correspond respectively to results for two different images from Deepfish dataset. The columns 1, 2 and 3 correspond respectively to results using : (a) Robust Teacher, (b) Unbiased Teacher, and (c) Active Teacher.

5.3 Experimental Results

5.3.1 Performance Analysis of Existing SSOD Methods

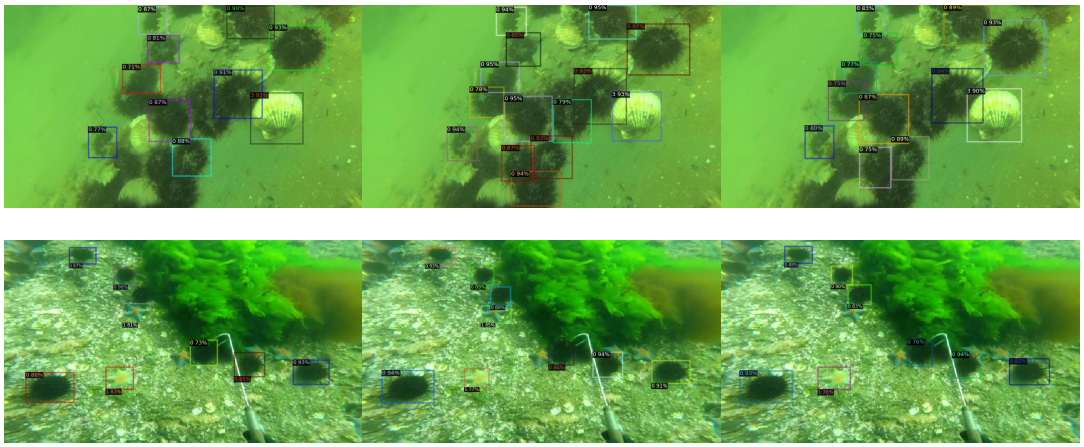
Fig.1 presents fish detection results obtained for different images from the validation set of Deepfish dataset. Color distortion, low contrast, blurred regions, and variations in fish appearances are noticed in these images. As observed, the different methods provide a good detection results with differences in performance. They detects the boundaries of fish with different forms and sizes (even small ones). Besides, a number of grouped fish are successfully separated as marked with their corresponding bounding box. However, we still have missing or wrong detections, and we have others with low accuracy. As shown in Fig.1-(b), the Unbiased Teacher outperforms the Active Teacher and Robust Teacher in number of correct detection and precision which can be explained by the fact that the Unbiased teacher uses the portion of the unlabeled dataset effectively to improve detection. However, Robust Teacher, being more focused on noise handling, and Active Teacher, being focused on selective labeling, may not make full use of the abundant unlabeled data as efficiently as Unbiased Teacher. For quantitative evaluation, results are resumed in Table.2, which are obtained using AP metrics.

Table 2: Detection results on Deepfish dataset with popular semi-supervised methods.

Methods	AP (%)	AP_S (%)	AP_M (%)
Supervised Faster-RCNN	56.10	21.20	46.70
Robust Teacher	58.85	24.58	49.05
Active Teacher	60.00	27.83	50.22
Unbiased Teacher	66.83	39.75	57.38

Results confirm that Unbiased Teacher outperforms the Robust and Active Teachers and the supervised Faster-RCNN. As an example, AP (%) equals 66.83 for Unbiased Teacher, while it is only 60 for Active Teacher, 58.85 for the Robust Teacher, and 56.10 for Faster-RCNN. The detection results for small, medium objects are improved using Unbiased Teacher compared to the other models. Besides, semi-supervised models can achieve baseline supervised performance (e.g., Faster R-CNN) with much less label expenditure. For instance, the supervised Faster R-CNN achieves 60% AP with 100% labeled data, while Active Teacher reaches similar performance with only 40% labeled data. Unbiased Teacher achieves superior performance, reaching 66.83% AP, as shown in Table.2. However, it is important to note that these semi-supervised methods do not reach the performance level of SOTA fully-supervised methods. The results of the SOTA fully-supervised methods will be provided in the appendix for comparison.

Fig.2 presents detection results of underwater animals in two different images from the validation set of UTDAC2020 dataset. The same as for Deepfish dataset, UTDAC2020 dataset suffers from low contrast, blur regions, and color distortion. As observed, Active Teacher, Unbiased Teacher, and Robust Teacher succeed in recognizing more than one category and detecting animals with different sizes and forms. However, detection is not optimal (missing detections). More detections and precision marked with bounding boxes, are obtained using Active Teacher and unbiased Teacher compared to Robust Teacher. Performances can be explained by the fact that Robust Teacher may focus on improving overall stability or robustness by dealing with noise



(a) Robust Teacher (b) Unbiased Teacher (c) Active Teacher

Figure 2: Rows 1 and 2 correspond respectively to results for two different images from UTDAC2020 dataset. The columns 1, 2 and 3 correspond respectively to results using : (a) Robust Teacher, (b) Unbiased Teacher, and (c) Active Teacher.

in the dataset, but that alone does not ensure better performance. However, the Unbiased Teacher and the Active Teacher focus on the ambiguous or poorly predicted instances in images and allocate more resources to learning these cases. Quantitative evaluation are provided in Table.3. Results attest that the Unbiased Teacher outperforms the other presented methods in terms of performance. The AP (%) for Unbiased Teacher is 44.22, compared to 43.86, 40.97, and 39.50 for Active Teacher, Robust Teacher, and supervised Faster-RCNN, respectively. Additionally, the Unbiased Teacher surpasses the baseline fully supervised Faster-RCNN which has 44% of AP (not reported in the Table.3).

Table 3: Detection results on UTDAC2020 dataset with popular semi-supervised methods.

Methods	AP (%)	AP_S (%)	AP_M (%)
Supervised Faster-RCNN	39.50	15.20	35.40
Robust Teacher	40.97	15.74	35.91
Active Teacher	43.86	15.96	38.97
Unbiased Teacher	44.22	17.92	38.50

In addition, an evaluation of performance per category, is given by Table.4. The Unbiased Teacher demonstrates the best overall performance, especially with Echinus and Holothurian, and it handles Scallop and Starfish detections better than others. Active Teacher is relatively consistent, particularly strong with Starfish detection, but not as effective for Holothurian. Robust Teacher consistently performs the worst, struggling the most with Holothurian (only 30.88%), and generally falling behind in all categories. Its results suggest that it may be less suited for this specific detection task. In this case, Unbiased Teacher offers the most balanced and effective solution across different marine species.

Table 4: Detection results on UTDAC2020 dataset per category with popular semi-supervised methods.

Methods	Echinus	Scallop	Starfish	Holothurian
Robust Teacher	43.75	40.00	49.25	30.88
Active Teacher	43.38	46.94	50.45	34.67
Unbiased Teacher	45.27	44.67	50.78	36.14

These results highlight the good potential of semi-supervised models when applied to underwater datasets. Additionally, they offer a promising alternative to supervised models, that rely on large amounts of labeled data which can be challenging to obtain in the context of underwater imagery. However, the detection process remains not optimal, with missed and wrong detections with low accuracy still observed in several images. To address this, we propose incorporating two key elements for underwater applications to semi-supervised models: UIE methods and contrastive learning, and evaluating their impact on the detection process. This will be the focus of the upcoming ablation study.

5.3.2 Ablation Study: UIE Methods

In this section, we applied different UIE methods to Deepfish and UTDAC2020 datasets. Both training and validation sets are enhanced by the same UIE technique. The UIE-1 adjusts the color distribution of the underwater image and uses the multi-scale fusion to improve the overall quality, enhancing the clarity and contrast of the image. UIE-2 restores natural colors, enhances contrast, and preserves the fine details, and UIE-3 improves the clarity of the image and brings out fine details that are lost in a hazy underwater environment. We investigate the impact of enhanced images through UIE methods on marine habitats detection. Table.5 and Ta-

Table 5: Detection results of semi-supervised methods with UIE for Deepfish dataset.

Methods	UIE	AP (%)	AP_S (%)	AP_M (%)
Robust Teacher	UIE-1	57.80 (-1.05)	26.05 (**+1.47**)	48.60 (-0.45)
Active Teacher	UIE-1	58.94 (-1.06)	29.58 (**+1.75**)	50.35 (**+0.13**)
Unbiased Teacher	UIE-1	66.58 (-0.25)	38.50 (-1.25)	57.06 (-0.32)
Robust Teacher	UIE-2	52.32 (-6.53)	17.55 (-7.03)	41.93 (-7.12)
Active Teacher	UIE-2	54.04 (-5.96)	19.41 (-8.42)	44.69 (-5.53)
Unbiased Teacher	UIE-2	63.28 (-3.55)	33.53 (-6.22)	53.04 (-4.34)
Robust Teacher	UIE-3	58.51 (-0.34)	24.43 (-0.15)	49.00 (-0.05)
Active Teacher	UIE-3	59.60 (-0.40)	28.74 (**+0.91**)	50.65 (**+0.43**)
Unbiased Teacher	UIE-3	66.66 (-0.17)	38.56 (-1.19)	57.16 (-0.22)

Table 6: Detection results of semi-supervised methods with UIE for UTDAC2020 dataset.

Methods	UIE	AP (%)	AP_S (%)	AP_M (%)
Robust Teacher	UIE-1	39.64 (-1.33)	14.66 (-1.08)	34.41 (-1.50)
Active Teacher	UIE-1	42.23 (-1.63)	14.36 (-1.60)	37.08 (-1.89)
Unbiased Teacher	UIE-1	42.92 (-1.30)	16.96 (-0.96)	37.35 (-1.15)
Robust Teacher	UIE-2	32.54 (-8.43)	12.32 (-3.42)	31.05 (-4.86)
Active Teacher	UIE-2	35.67 (-8.19)	14.59 (-1.37)	34.49 (-4.48)
Unbiased Teacher	UIE-2	35.91 (-8.31)	13.68 (-4.24)	33.80 (-4.70)
Robust Teacher	UIE-3	40.37 (-0.60)	16.08 (**+0.34**)	35.26 (-0.65)
Active Teacher	UIE-3	42.72 (-1.14)	16.10 (**+0.14**)	37.38 (-1.59)
Unbiased Teacher	UIE-3	44.00 (-0.22)	17.50 (-0.42)	38.25 (-0.25)

ble.6 show the AP values obtained by applying semi-supervised models to the enhanced DeepFish and UTDAC2020 datasets, respectively. The values in parentheses represent the improvement compared to the performance without the UIE module. The AP results in Table.5 and Table.6 attest the non linearity correlation between of the image enhancement and the accuracy of the object detection model. Although, image enhancement methods, performed well in the visual sense. For Deepfish and UTDAC2020 datasets, they do not achieve better detection accuracy with Active Teacher, Robust Teacher, and Unbiased Teacher. The accuracy of semi-supervised models declines after applying underwater image enhancement, compared to their original performance. For example, the original performance of Robust Teacher on deepfish dataset is identified with $AP(\%)$ equals 58.85, while it is decreased to 57.80, 52.32, and 58.51 when applying UIE-1, UIE-2, and UIE-3, respectively.

Many reasons can explain the inconsistency between enhancing the image quality and the detection performance of semi-supervised model; the absence of Ground Truth images for UIE methods make the enhanced image not necessarily better than the original image, besides, the optimization objective of UIE method is different from that of an underwater object detection model. The two objectives are not aligned with one another. The purpose of UIE is only to ameliorate the human visual senses of an image, while the detection model aims to locate underwater targets. Therefore, it is not practical to use UIE methods as a pre-processing step for underwater object detection only based on quality metrics. More efforts are needed to ensure more effective methods for quality assessment.

5.3.3 Ablation Study: Contrastive Learning

In further experiments, we integrate contrastive learning with the teacher-student architecture employed in Active Teacher, Robust Teacher, and Unbiased Teacher, without applying any UIE techniques. The AP values for the DeepFish and UTDAC2020 datasets using contrastive semi-supervised models are summarized in Table 7. The values in parentheses represent the improvement compared to the performance without the contrastive learning. These results illustrate the contribution of contrastive learning in improving certain detection results. As illustrated in Table.7, Deepfish and UTDAC2020 detection results are slightly ameliorated. Especially, the average of detection for small underwater targets is more refined as noticed for Unbiased Teacher and Active Teacher. As an example, for Deepfish dataset, AP_s equals 30.21% for Active Teacher with the integration of contrastive branch, compared to only 27.83% without it, resulting in a 2.38% improvement. Although the improvement provided by contrastive learning is not yet significant, we believe that with further research and more sophisticated integration techniques like the work in (Seo et al., 2022; Wu et al., 2022), contrastive learning has the potential to enhance detection results, particularly for small and medium marine habitats. These

Table 7: Detection results of semi-supervised methods with incorporated contrastive learning.

Datasets	Methods	AP (%)	AP_S (%)	AP_M (%)
Deepfish	Robust Teacher	58.79 (-0.06)	23.33 (-1.25)	48.85 (-0.20)
	Active Teacher	60.19 (**+0.19**)	30.21 (**+2.38**)	51.27 (**+1.05**)
	Unbiased Teacher	66.93 (**+0.10**)	41.66 (**+1.91**)	57.20 (-0.18)
UTDAC 2020	Robust Teacher	41.04 (**+0.07**)	15.74 ($\pm$0.00)	35.67 (-0.24)
	Active Teacher	43.43 (-0.43)	15.54 (-0.42)	38.63 (-0.34)
	Unbiased Teacher	44.30 (**+0.08**)	18.67 (**+0.75**)	38.20 (-0.30)

advanced techniques require specifically designed algorithms tailored for semi-supervised settings. Incorporating them into this study would have necessitated significant additional development, which falls beyond the scope of our current objectives. Therefore, we have left the exploration of such techniques for future work.

6 CONCLUSIONS

In this paper, we proposed a comprehension analysis of marine habitats detection results, performed using different semi-supervised methods. The latter represent an alternative to supervised ones, to deal with the presence of limited labeled data, which is the case for underwater datasets. Results encompass a focus on Active Teacher, Unbiased Teacher, and Robust Teacher as semi-supervised models, applied to Deepfish and UTDAC2020 datasets. In this work, we proposed UIE methods to enhance the image quality and used these enhanced images as input for semi-supervised models. In addition, we introduced a new contrastive branch to study its impact on marine habitats detection. Qualitative and quantitative evaluations are attested through many experiments. They both demonstrate the significant performance of semi-supervised models in detecting underwater images. On the other hand, we conclude that enhanced images do not obligatory improve detection results, while the integration of contrastive branch can result in refined detection, where small and medium underwater targets are more located. In future work, we aim to explore two key directions: first, improving contrastive learning to enhance the feature representation; and second, directly integrating the Underwater Image Enhancement module as a domain-specific augmentation technique.

ACKNOWLEDGEMENTS

This work is fully funded by the project ROV-Chasseur (ANR-21-ASRO-0003) of the French National Research Agency (ANR).

REFERENCES

Alaba, S. Y., Shah, C., Nabi, M., Ball, J. E., Moorhead, R., Han, D., Prior, J., Campbell, M. D., and Wallace, F. (2023). Semi-supervised learning for fish species recognition. In *Ocean Sensing and Monitoring XV*, volume 12543, pages 247–254. SPIE.

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). Enhancing underwater images and videos by fusion. In *2012 IEEE conference on computer vision and pattern recognition*, pages 81–88. IEEE.

Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., and Bekaert, P. (2017). Color balance and fusion for underwater image enhancement. *IEEE Transactions on image processing*, 27(1):379–393.

Bartyzel, K. (2016). Adaptive kuwahara filter. *Signal, image and video processing*, 10:663–670.

Chen, L., Liu, Z., Tong, L., Jiang, Z., Wang, S., Dong, J., and Zhou, H. (2020). Underwater object detection using invert multi-class adaboost with deep learning. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Chen, X. and Chen, H. (2010). A novel color edge detection algorithm in rgb color space. In *IEEE 10th International Conference On Signal Processing Proceedings*, pages 793–796. IEEE.

Cho, J. W., Kim, D.-J., Jung, Y., and Kweon, I. S. (2022). Mcdal: Maximum classifier discrepancy for active learning. *IEEE transactions on neural networks and learning systems*.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., and Le, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *CVF conference on computer vision and pattern recognition*, pages 113–123.

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Han, F., Yao, J., Zhu, H., Wang, C., et al. (2020). Underwater image processing and object detection based on deep cnn method. *Journal of Sensors*, 2020.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 770–778.

Islam, M. J., Xia, Y., and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception.

IEEE Robotics and Automation Letters, 5(2):3227–3234.

Jahanbakht, M., Azghadi, M. R., and Waltham, N. J. (2023). Semi-supervised and weakly-supervised deep neural networks and dataset for fish detection in turbid underwater videos. *Ecological Informatics*, 78:102303.

Jeong, J., Verma, V., Hyun, M., Kannala, J., and Kwak, N. (2021). Interpolation-based semi-supervised learning for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11602–11611.

Jiang, B., Luo, R., Mao, J., Xiao, T., and Jiang, Y. (2018). Acquisition of localization confidence for accurate object detection. In *European conference on computer vision (ECCV)*, pages 784–799.

Lau, P. Y. and Lai, S. C. (2021). Localizing fish in highly turbid underwater images. In *International Workshop on Advanced Imaging Technology (IWAIT) 2021*, volume 11766, pages 294–299. SPIE.

Li, S., Liu, J., Shen, W., Sun, J., and Tan, C. (2023). Robust teacher: Self-correcting pseudo-label-guided semi-supervised learning for object detection. *Computer Vision and Image Understanding*, 235:103788.

Liu, Y.-C., Ma, C.-Y., He, Z., Kuo, C.-W., Chen, K., Zhang, P., Wu, B., Kira, Z., and Vajda, P. (2021). Unbiased teacher for semi-supervised object detection. *International Conference on Learning Representations (ICLR)*.

Mi, P., Lin, J., Zhou, Y., Shen, Y., Luo, G., Sun, X., Cao, L., Fu, R., Xu, Q., and Ji, R. (2022). Active teacher for semi-supervised object detection. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Pan, T.-S., Huang, H.-C., Lee, J.-C., and Chen, C.-H. (2021). Multi-scale resnet for real-time underwater object detection. *Signal, Image and Video Processing*, 15:941–949.

Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing*, 32:3066–3079.

Popat, S. K., Deshmukh, P. B., and Metre, V. A. (2017). Hierarchical document clustering based on cosine similarity measure. In *International Conference on Intelligent Systems and Information Management (ICISIM)*, pages 153–159. IEEE.

Ren, B., Feng, J., Wei, Y., and Huang, Y. (2022). Underwater target detection algorithm based on improved yolov5. *Advances in Engineering Technology Research*, 1(3):713–713.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28.

Saleh, A., Laradji, I. H., Konovalov, D. A., Bradley, M., Vazquez, D., and Sheaves, M. (2020). A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Scientific Reports*, 10(1):14671.

Sarkar, P., De, S., and Gurung, S. (2022). A survey on underwater object detection. *Intelligence Enabled Research: DoSIER*, 1029:91–104.

Seo, J., Bae, W., Sutherland, D. J., Noh, J., and Kim, D. (2022). Object discovery via contrastive learning for weakly supervised object detection. In *European Conference on Computer Vision*, pages 312–329. Springer.

Sohn, K., Zhang, Z., Li, C.-L., Zhang, H., Lee, C.-Y., and Pfister, T. (2020). A simple semi-supervised learning framework for object detection. *AAAI Conference on Artificial Intelligence*.

Song, P., Li, P., Dai, L., Wang, T., and Chen, Z. (2023). Boosting r-cnn: Reweighting r-cnn samples by rpn's error for underwater object detection. *Neurocomputing*, 530:150–164.

Song, W., Wang, Y., Huang, D., Liotta, A., and Perra, C. (2020). Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Transactions on Broadcasting*, 66(1):153–169.

Srividhya, K. and Ramya, M. (2017). Accurate object recognition in the underwater images using learning algorithms and texture features. *Multimedia Tools and Applications*, 76:25679–25695.

Tarvainen, A. and Valpola, H. (2017). Weight-averaged consistency targets improve semi-supervised deep learning results. *Neural Information Processing Systems (NeurIPS)*.

Wang, Y., Liu, Z., and Lian, S. (2023). Semi-supervised object detection: A survey on recent research and progress. *arXiv:2306.14106*.

Wu, W., Chang, H., Zheng, Y., Li, Z., Chen, Z., and Zhang, Z. (2022). Contrastive learning-based robust object detection under smoky conditions. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 4294–4301.

Xie, Q., Dai, Z., Hovy, E., Luong, T., and Le, Q. (2020). Unsupervised data augmentation for consistency training. *Advances in neural information processing systems*, 33:6256–6268.

Xu, S., Zhang, M., Song, W., Mei, H., He, Q., and Liotta, A. (2023). A systematic review and analysis of deep learning-based underwater object detection. *Neurocomputing*, 527:204–232.

Yu, H. (2020). Research progresson object detection and tracking techniques utilization in aquaculture: a review. *Journal of Dalian Ocean University*, 35(6):793–804.

Zhang, F., Pan, T., and Wang, B. (2022a). Semi-supervised object detection with adaptive class-rebalancing self-training. In *AAAI conference on artificial intelligence*, volume 36, pages 3252–3261.

Zhang, Y., Zhang, X., Li, J., Qiu, R. C., Xu, H., and Tian, Q. (2022b). Semi-supervised contrastive learning with similarity co-calibration. *IEEE Transactions on Multimedia*, 25:1749–1759.

Zhou, Y., Hu, D., Li, C., and He, W. (2023). Uwyolox: An underwater object detection framework based on image enhancement and semi-supervised learning. In *International Conference on Neural Computing for Advanced Applications*, pages 32–45. Springer.

Zurowietz, M. and Nattkemper, T. W. (2020). Unsupervised knowledge transfer for object detection in marine environmental monitoring and exploration. *IEEE Access*, 8:143558–143568.